Chapter 1:

Introduction

Valuation of health states by the general population is being attempted recently. Most of these studies have been in fully literate societies in the economically developed countries. For example, in Canada by Sacckett and Torrance (1978), UK (Gudex and others, 1996). Because population-based empirical assessments of health states are extremely limited, new surveys are needed. A study was conducted in Andhra Pradesh, by the Institute of Health Systems (IHS) to measure peoples preferences about various health states. To the best of our knowledge, this is the first community based health state valuation study in the developing world. Two distinct sources of assessment were made. First arm of the study consisted of a series of multi method health state valuation workshops for educated persons from different backgrounds. Participants of these workshops, valued health states using more than one procedure including, card sort, visual analogue scale, time trade-off, and person trade-off methods. Second arm of the study involved measurement of valuations given by general population drawn from a rural area. This was done by household surveys in a selected village in Ranga Reddy district. Respondents in the survey were requested to give their valuations using card sort followed by visual analogue scales. The study was conducted in the year 1999 and was funded by the WHO - Global Program on Evidence for Policy (GPE). Broad goals of this study were;

1. To strengthen the methodological foundation for population based measurement of health state weights. In addition valuation protocols must be adapted for use in partly literate populations like that in India.
2. To measure local preferences for health states to be used for estimation of national and state burden of disease in India.
3. To understand the nature health state valuation function in people's mind and its implications for summary measures of population health status.

This monograph is organised into seven chapters including this introduction. Selection of index health states (conditions) for the study. This chapter introduces the subject of health state valuation, reviews current literature and outlines various methods of measurement. Chapters 2 - 5 describe methodology and organisational details of the study. Chapter 6 analyses data from the two arms of the study, to assess reliability and validity of the health state valuation measurements. Analysis of test retest data also provides important

insights into the nature of health state valuation function in people's mind. Results of these analyses are presented in chapter 6. Chapter - 7 presents results from the community survey of health state valuations and summarises findings of the study. The monograph includes a large number of appendices containing various methodological and organisation details. We hope that these details will help researchers to easily implement health state valuation studies and contribute to reliability in study methodology.

Health state or disability weights for computation of disease burden:

The key step in construction of synthetic measures of health status is assignment of weights to different health states. This methodological significance in construction of synthetic measures need to be distinguished from the quantitative impact, alternate sets of disability weights have on total disease burden estimate. Murray and Lopez (1996 p288) observed that rank order of diseases and injuries was insensitive to alternate set of disability weights. But relative size of disability to mortality components of disease burden changed. Using a set of disability weights sensitive to minor and trivial illnesses and small deviations from perfect health state decreased the disability adjusted life expectancy. Allen and others (1989) felt that ordinal ranking of cost per QALY for different interventions would not change with appreciable changes in corresponding index of health status. They observed that life saving procedures would always tend to score better than palliative or pain relieving measures which in turn would show lower cost to effectiveness ratio than expensive continuing therapy. However, there are other compelling reasons to attach importance to assignment of disability weights. Firstly, robustness of disease burden estimates or cost-effectiveness ratios to alternate set of disability weights is a feature of the current epidemiological state. As mortality continues to decline and prevalence of degenerative diseases rise further, the importance of disability weight for these results will increase. Secondly alternate sets of health state weights will certainly affect composition of the disability component of disease burden. This may, in certain circumstances, be an important input to policy analysis. For example, suppose mortality in a country has declined to a level close to our understanding of the biological potential of longevity. Then it would make sense

to analyse the composition of disability as such to prioritise between interventions seeking to improve health related quality of life. Last but not the least, the whole purpose of seeking out synthetic measures of health status would be defeated if adequate attention is not paid to component subjects of synthesis. More over, most health related quality of life measurements have taken place in the industrialised and economically developed countries. An important concern has been if the health state weights are robust across various cultural settings. Health related quality of life is now sought to be defined by restricting to domains of functioning that are universally, most essential to one's ability to pursue valued life goals (Shumaker and Naughton, 1995). Thus local measurement of disability weights is important from two perspectives, namely (a) sensitivity of national disease burden estimates to locally measured disability weights as opposed to use of global disability weights and (b) understanding health status weights across cultures.

Note that disability weights used in computation of DALYs is the complement of health state weights or quality adjustment weights. Estimating one gives the other by simple arithmetic manipulation (disability weight = 1 - quality adjustment weight). These terms used here interchangeably.

Interpretations of disability weight:

Theoretical interpretations about the object of measurement has some bearing on the methodological path leading up to health status weights. Three different interpretations of have been made in the literature, namely; (a) individual preference, (b) descriptive measure of health state and (c) social preference weight attached to the health state. Culyer (1989) distinguished between "welfarist[1]" and "extra-welfarist" approaches, among economists, and discussed its implication for the health sector. The "welfarist" approach is to view individual preferences (utility) as the source of all social welfare[2]. Health state or disability weight is viewed to represent individual preference for different health states. Since health outcomes at a personal level is characterised by uncertainty, von Neuman and Morgenstern's (VNM)

[1] According to Culyer, the "welfarist" label was given by Amartya Sen (1977).
[2] There are important and unresolved issues regarding aggregation of individual preferences for quantification of social welfare. The impossibility of impossibility of aggregating cardinal measures of utility that would also satisfy certain reasonable assumptions. was shown by Arrow.

expected utility theory[3] is applied. Thus health state weights are viewed as VNM utility. Viewing health state weight as a measure of personal utility means that an equivalence with utility of other goods and services is straight forward. Advocates of personal utility interpretations may not make use of this equivalence in deference to strong emotional responses against valuation of human life in money terms. Nonetheless, the equivalence exists. From the Welfarist point of view, there is no theoretical basis for use rating scales for health state valuation, since these do not require the valuer to choose between two alternatives. The key theoretical underpinning of measurement of preference is ordering of alternatives by the individual. Since, rating scales simply ask the valuer to assign a weight in a scale without any reference to other health states, Welfarist's would argue, the resultant measurement can not be interpreted as a measure of utility. Welfarists may, however, tolerate visual analogue scaling as practical measurement device and seek to postulate empirically observed relationships between visual scaling and standard gamble or other trade-off methods.

The extra-welfarist approach of Amartya Sen (1979), Culyer argues, admits non-utility information about individuals into the process of comparing social states. An extra-welfarist approach allows for the concept of need as the basis of social welfare. The concept of need implies the existence of a goal which is considered reasonable. Extra-welfarists view health itself as a descriptive characteristic of people and as the principal maximand (out come to be maximised) of health services. The health state weight describes this output i.e. the health related quality of life (Bleichrodt, 1997). This later view allows for measurement by rating scales and magnitude estimation tools like the visual analogue scales, in addition to utility theory based measurement tools.

The third interpretation, that quality adjustment weights are values under a social welfare function is proposed by Nord (1995). Patrick et al (1973) also recognised that quality adjustment weight derived from the equivalence (person trade-off) method would represent valuation under some social welfare function.

---

[3] von Neumann and Morgenstern (1944) proposed the expected utility theory. This theory is described in any text books on micro economic theory. One good exposition is in Mas-Collel, Whinstone and Green (1995).

Measurement strategy:

Torrance (1986) gives an overview of measurement of health state utilities for economic appraisal. In a four part series, Froberg and Kane (1989 I-IV) have summarised methodology for measurement of health state weights. The first article (Froberg and Kane, 1989-I) deals with measurement strategy, which refers to the overall structure for posing questions to respondents and corresponding method of data analysis. They divide extant measurement strategies into (a) holistic and (b) decomposed approaches. Torrance (1986) viewed these as alternative description of health states and used a similar classification, labelling the decomposed approach as health state classification system. In holistic approach each of the full range of health states is described as a whole including all its attributes. The respondent is presented with the description for all health states one after the other and asked to assign values. Thus the respondent has to judge each health state as a whole and all health states in the scale. As a result the procedure becomes cognitively demanding for the respondent. In decomposed designs the respondent does not have to value all health states in the profile. Decomposition may be explicit or achieved by statistical modelling. For statistically inferred decomposition the respondent is presented with the multiattribute description of a health state as in case of holistic approach. But only a few health states are presented to a single respondent, thus reducing cognitive overload. An algebraic model of multiattribute health states is constructed using statistical inferences from respondent evaluations. Explicit decomposition for health state measurement is rooted in multiattribute utility theory (Torrance 1982, 1986). Here the respondent is asked to evaluate each dimension of health separately, there by keeping it cognitively simple. Froberg and Kane (1989-I) recommend the statistically inferred strategy in view of its simplicity for respondents.

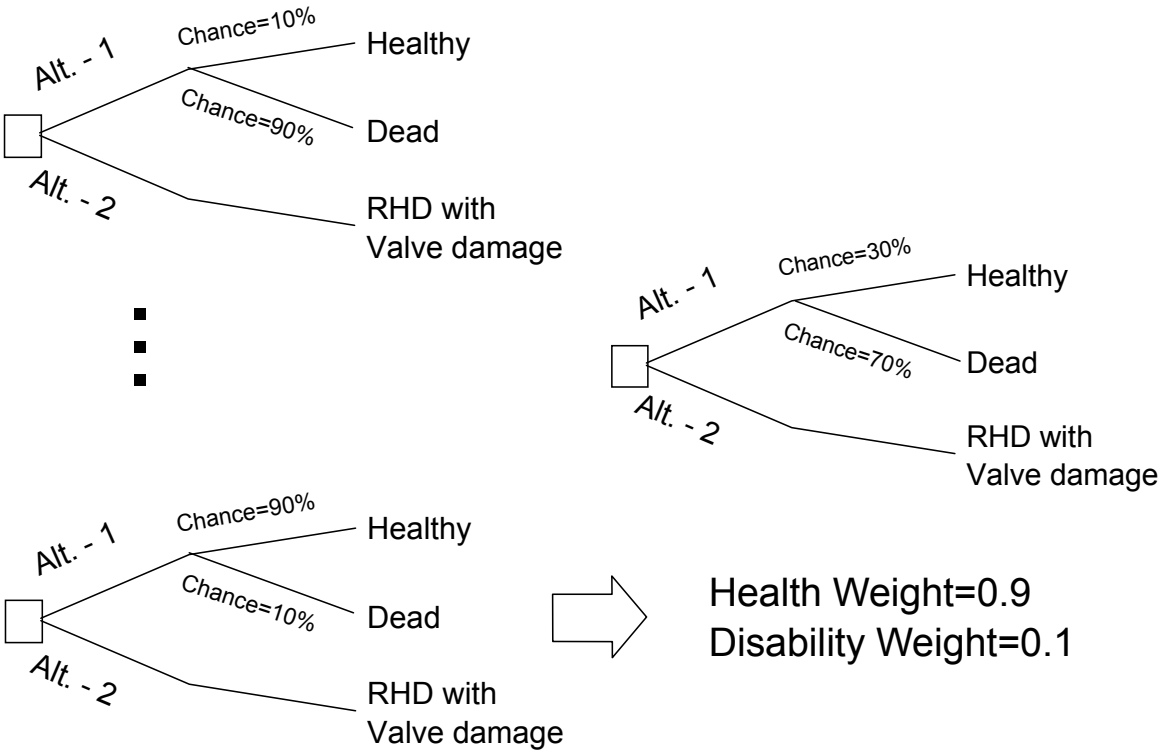
Scaling methods

In the second article Froberg and Kane (1989-II) list six scaling methods. These are; (a) standard gamble, (b) time trade off, (c) rating scale, (d) magnitude estimation, (e) equivalence or person trade off, and (f) willingness to pay. Rating scales and magnitude

estimation methods are psychometric in nature. Standard gamble, time trade off and willingness to pay are all preference based. Person trade off is preference based and has psychometric origins as well. A brief description of each of these methods is presented below.

Standard gamble:

Figure-1.1: An example of successive personal decision alternatives with uncertain outcomes (gambles) for valuation of the health state rheumatic heart disease (RHD) with valve disease.

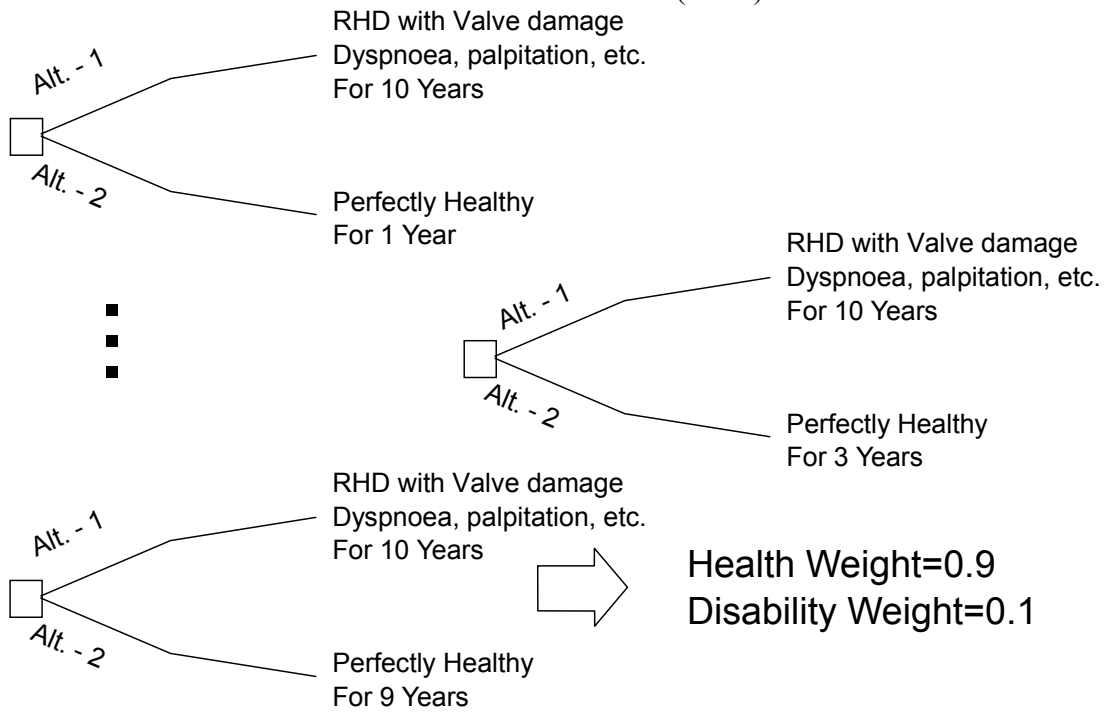


The standard gamble measures presents a random prospect (gamble), consisting of a best and a worst outcome, and the alternative of a certain prospect intermediate in desirability between the best and worst outcomes. The random prospect is completely defined by probability *p* of one of the two possible events. The other event will have a probability (*1-p*). For example the best outcome in the random prospect can be "perfectly healthy" and the worst outcome can be "dead". The certain prospect of intermediate desirability will then be a specific health state for example living with a particular disease state. To obtain quality adjusted weights the random prospect is defined by assigning probability *p* to the "perfectly healthy" outcome. The two alternative prospects defined by choice of disease state for the certain arm and *p* for the random arm are presented to a respondent with some initial value of

p. The value of *p* is then varied till the respondent is indifferent between the certain prospect and the random prospect. The *p* that satisfies this condition is taken as the quality adjustment weight for the disease condition defined in the certain arm. Disability weight for this health state is given by *(1-p)*. Instead we can define the random prospect by assigning probability *q* to the worst outcome (death in this set up). In that case the value of *q* satisfying the indifference condition is the disability weight. The standard gamble was proposed by von Neuman and Morgenstern (1944) as a tool to measure expected utility. The difficulty with this method is that it is not easily understood by many respondents.

Time trade off:

Figure-1.2: An example of successive personal time trade-off alternatives for valuation of the health state rheumatic heart disease (RHD) with valve disease.



Time trade-off was designed by Torrance et al (1973) for health status measurement, as a simpler and cognitively less demanding alternative to the standard gamble. Here the valuer's preferences for health states is assessed indirectly through the time (number of years, months, days, hours) (s)he wants to trade in to lead a completely healthy life, as against life with a particular less-than-perfect health state. The subject is offered two alternatives. Alternative 1: state *i* for time *t* (local life expectancy of an individual with the chronic

condition) followed by death. Alternative 2 is a perfectly healthy state for time *x,* where *x* is less than *t*. Time *x* is varied until the respondent is indifferent between the two alternatives, at which point the required preference value for state *i* is given by $h_i = \frac{x}{t}$.

Rating scales:

Rating scale consist of a range of values with clearly defined end points (anchors). For holistic health status measurement "perfect health" and "death" act as natural anchors. The range of values could be continuous or discrete. The visual analogue scale (VAS) consisting of a graduated line segment, one end labelled as death and the other labelled as perfect health is a continuous rating scale. Another form of continuous rating scale uses adjectival labels to describe intermediate points of a line anchored at both ends (for example see fig 4.4b at page 34 in Streiner and Norman, 1995). Except for the intermediate labels and smaller line length, they resemble the VAS. Rating scales using equally appearing intervals i.e. discrete points along the scale are called category rating scales (for example see fig 4.4a at page 34 in Streiner and Norman, 1995). Streiner and Norman (1995) describe rating scales as direct estimation methods (chapter 4 page 32-39) and category or continuous rating scales are described as adjectival scales. Rating scales are most frequently used to measure health state weights. Specter (1992) describes theoretical basis and practical steps in construction of rating scales for general psychometric use. Streiner and Norman (1995) provide similar description for construction of rating scales in the field of health status measurement.

The visual analogue scales are simple in construction. But respondents may not always agree. For example Streiner and Norman (1995) cite a study by Huskisson (1974) in which 7% of patients were unable to complete a VAS against 3% for category rating scale. Bosi Refaz et al (1990) found that illiterate respondents had more difficulty with VAS compared to category rating. In a comparative study of VAS and person trade off method, Nord (1991) asked his respondents to describe the meaning they attached to points in VAS chosen by them. Both ends of the VAS was anchored by worst and best imaginable health state, respectively. Sixty seven (out of a total of 105) respondents answered this question. Nineteen persons said that they viewed it as a percentage of best imaginable health state. Eleven did not mean anything and the remaining 37 did not answer the question directly but

described specific dimensions of given condition taken into consideration by them. Nord observed that subjects were expressing strength of preferences, through the VAS, in addition to ranking of health states. Person trade off implied by VAS were lower than ratios obtained from directly asked person trade off questions. With this finding coupled with the small size (19) of respondents reporting a conscious attempt to relate given state in percentage terms to one end of the scale, Nord concluded that one should not put too much emphasis on the numerical values obtained from VAS. But such an interpretation is flawed by many limitations of this study. Firstly, Nord appears to be denominating the 19 persons consciously trying to assign a ratio number with the 67 persons who gave some response to their question on meaning of valuations. More than half (37) of them did not give a direct answer to this question. So appropriate denominator to appreciate the relative size of the 19 consciously trying ratio raters would be the 30 persons who gave a direct answer to this question. Secondly Nord uses person trade off as a criterion, to compare results from VAS. The person trade off methodology is itself very sensitive to sample size, framing and start point bias found subsequently by Nord (1995).

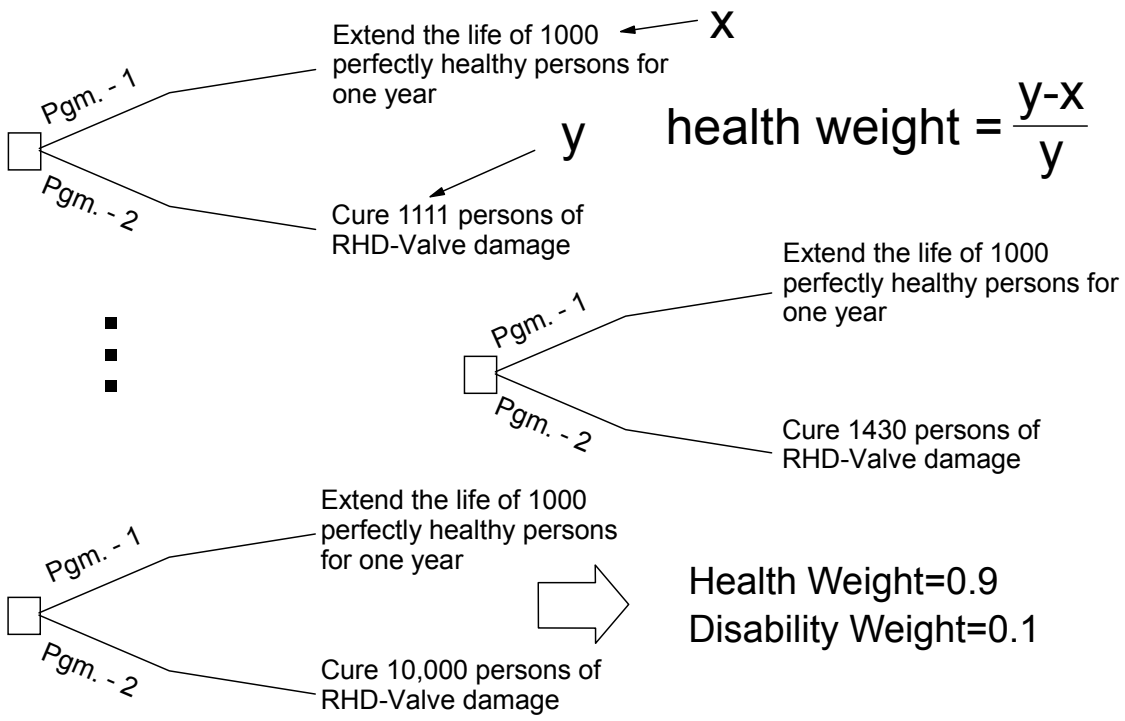
Magnitude estimation:

In magnitude estimation a reference state is identified and described. A numerical value (numerical estimation) or a line segment of certain length (line production) is assigned to the reference state. Respondents are presented with the health state of interest and asked to assess how it relates to the reference state. More specifically, how many times better or worse is the given health state in comparison to the reference state. They then assign a number proportional to the reference number (numerical estimation) or produce a line (line production) to show how it relates to the reference state. A higher number or larger length of line would mean better health state and a smaller number or line would mean worse health state. Calibration of respondents ability to reproduce magnitude estimates or at the least a practice session to orient them towards magnitude estimation is usually called for. Apart from the brief description in Froberg and Kane (1989b), McDowell and Newell (1987, 1996), magnitude scaling method is described by Lodge (1981).

Person trade off (PTO) or Equivalence:

Patrick et al (1973) adapted the method of adjustment or equivalent stimuli in psychometrics (Guilford, 1954; Torgerson, 1958) and devised the equivalence method for health status measurement. Froberg and Kane (1989b) opine that the person trade off (PTO) method is conceptually similar to magnitude scaling. In recent literature the same is referred to as person trade off method (Nord 1992, 1995). Here respondents are presented with two groups of persons distinguished by their health state and a constraint to the effect that only one of the two groups can be helped. One of the two groups is characterised by a reference health state (standard group) for example people with perfect health. The other group (evaluation group) is characterised by a specific health state of interest. The standard group with perfect health constitute the standard stimuli and number of persons in the evaluation group provides the variable comparison stimuli. Respondents are asked to choose between the two groups who should receive help. Number of persons in the evaluation group is varied till the respondent is indifferent between the two. Miles (1977) compared the person trade off method with category ratings. Difference in health states, between the two methods was not significant. Patrick and colleagues reported that this method was too complex. The unrealistic assumptions and the emotive nature of the task offended some judges. Similar observations were made by Nord (1995). In addition, Nord (1995) noted that the techniques need fairly large groups of respondents to keep measurement error within limits, is susceptible to start point bias, and is sensitive to question framing. To over come influence of question framing, respondents should be induced to arrive at an reflective equilibrium. To achieve this, Nord (1995) suggested a multistep procedure, through which respondents are presented all relevant arguments and to reconsider initial responses in the light of such arguments. A problem with this solution is; what arguments are considered relevant? For example Froberg and Kane (1989c) observed that “when an interviewer takes an active role the potential for influencing the rater is increased”. For purposes of population health status measurement and resource allocation, this method is assumed to be valid by construction, since it asks respondents to weigh the claims of one group of persons with respect to another, distinguished only by their health states (Patrick et al, 1973; Nord 1991, 1995; Murray 1996). Knowledge of its reliability is limited since the method has not been used widely enough (Froberg and Kane, 1989d).

Figure-1.3: An example of successive PTO-2 social decision alternatives for valuation of the health state rheumatic heart disease (RHD) with valve disease.



Murray (1996) used the PTO method to derive disability weights for the fifth revision of GBD estimates. A good deal of effort was made to minimise usual problems associated with this method. To induce deliberation and reflection two PTOs were designed using different situational characteristics. In the first method (PTO1) respondents trade life extension for persons with the health state of interest with life extension for healthy persons. In the second method (PTO2) respondents trade improving health related quality of life to complete well being with life extension for perfectly healthy persons. In general respondents assign a lower disability weight to the same health state in PTO1 compared to PTO2. Once the respondent completed the two PTOs for a given condition, the results are fed back to him while pointing out inconsistency if any between them. The respondent is then encouraged to revise the estimates to resolve the inconsistency. After evaluating all of the 22 indicator conditions, respondents are asked to give ordinal rank to these indicator conditions. This ordinal ranking is compared with the ranking implied by the reconciled PTO mentioned above. Any inconsistency is fed back to the respondent with instructions to reconcile them. Group discussions are encouraged at different stages to facilitate a consensus. Details of this protocol (GBD PTO protocol) are described in Murray (1996, pages 35-41 and appendix 1 at pages 90-98). A major deficiency that persisted is lack of standard description of health

states. Since most participants were chosen from health professions, it was assumed that they would have a shared understanding of the health states in question. Interactions between interviewers and respondents as well as the group discussions were not fully structured.

Willingness to pay:

Willingness to pay is a straightforward application of welfarist notion of health status as a personal utility similar to other goods and services in the economy. A persons willingness to pay for curing or avoiding impairment, disability and handicaps associated with a health state is taken as a measure of quality adjustment weight for that health state. Willingness to pay to avoid the risk of death can be used as a denominator to derive quality adjustment weights in the range of [0,1]. Apart from ethical objections concerning welfarist approach to health and economic valuation of human life, measurement of willingness to pay is complicated by lack of a perfectly (or nearly perfect) competitive market in the health sector. Various alternatives like the contingent valuation method, cost of illness method are used to get around this problem. A comprehensive description of these methods can be found in Tolley, Kenkel and Fabian (1994).

Controlling context effects:

Froberg and Kane (1989-III) in their third article review observations about the effect of respondent characteristics and other contextual characteristics on health status measurement. They group these factors into three clusters, (a) differences among population (respondent characteristics); (b) inconsistencies due to the nature of human judgement process (framing); and (c) inconsistency due to situation specific variables (situational differences). They observe that respondent characteristics do not have any significant effect on valuation and hence can be ignored. Framing effects can be reduced by presenting the problem in more than one way and asking the rater to reconcile inconsistencies. This is same

as the multistep approach (reflective equilibrium) suggested by Nord (1995). The solution to situational differences proposed by Froberg and Kane (1989-III) is to standardise them.

Psychometric instruments for general health status measurement:

A number of instruments are now available for health status measurement. McDowell and Newell (1987, 1996) provide an overview of these and distinguish two general classes, namely (a) instruments for measurement of general health status and (b) instruments designed for specific dimensions of health for example physical disability, psychological well being, pain etceteras. Our interest here is on the first category. McDowell and Newell (1996, chapter-9, pages 380-492) list 21 instruments for measurement of general health status. Eighteen of these instruments, called general health profiles, generate a profile of scores in different dimensions included in the instrument. Another three allow computation of a single index from out of the scores in component dimensions. These are called health indices. We will briefly describe four of these instrument, namely; (a) the sickness impact profile (SIP); (b) Short-form-36 (SF-36) health survey; (c) EuroQol, and (d) the quality of well being scale (QWB). The first two give a general health profile only. The last two produce general health indices in addition to profiles.

The sickness impact profile (SIP) seeks to measure changes in a person's behaviour on account of illness. Scoring is done along 12 categories or sub dimension. Respondent behaviour in each category is assessed by a set of questions graded according to severity or intensity along the sub dimension. There are altogether 136 such graded questions. The 12 sub dimensions can be grouped into (a) physical health consisting of ambulation, mobility and body care; (b) psycho social health consisting of social interaction, alertness, emotional state and communication; and (c) five independent categories, namely (i) sleep and rest, (ii) eating, (iii) work, (iv) home management and (v) recreation. Item weights were arrived at from more than 100 judges with equal appearing interval scaling procedures. The profile can

either be self administered of by an interviewer. It takes about 20 to 30 minutes to complete the questionnaire. The scale was developed by Bergner and others (1976a, 1976b, 1981).

The SF-36 instrument measures eight dimensions, namely (a) physical functioning, (b) role limitations due to physical problems, (c) pain, (d) social functioning, (e) general mental health, (f) role limitations due to emotional problems, (g) vitality, energy or fatigue and (h) general health perceptions. Physical functioning and role limitation due to physical problems can be viewed as one dimension. Similarly (d), (f) and (g) can be viewed as social functioning. Thus the dimensions covered in this instruments can summarised as (a) physical function status, (b) social function, (c) psychological well being and (d) pain. Each dimension is assessed by category rating of multiple items which are themselves graded by severity or intensity. The form uses preceding one month as the time frame for all questions. Alternative forms using shorter time frames for acute conditions have also been used. Reliability and criterion validity (using common sense criteria like ability to work, symptoms, etc.) have been found to be fairly high (McDowell and Newell, 1996). It was developed out of the health insurance experiment (HIE) and medical outcomes study (MOS) conducted by the RAND corporation in USA. The SF-36 instrument has been described by Ware and Sherbourne (1992), Ware and others (1993), McHorney and others (1993, 1994), Anderson and others (1993).

The EuroQol is a summated rating scale consisting of five dimensions. These are mobility, self care, usual activities, pain or discomfort and anxiety or depression. Each dimension is rated by a three point category rating scale. Weights for computation of composite index were developed by using valuation of 10 core health states on visual analogue scales. The instrument has four parts as follows: (a) description of patients own health (page 2) along five dimensions; (b) overall rating of own health using a visual analogue scale (page 3); (c) valuation of a standard set of health states (pages 4-7); and (d) background information about the respondent (pages 8-9). Parts (a) and (b) are required to collect data on health related quality of life. A general health index can be computed using weights derived by the EuroQol team using responses from the valuation part (pages 4-7). For local weights these parts have to be implemented as well. The EuroQol instrument is described by the EuroQol group (1990), Brooks and the EuroQol group (1996), and

McDowell and Newell (1987, 1996). The full instrument is reproduced in Shumaker and Berzon (1995 appendix -2).

The quality of well being (QWB) scale is a summated rating scale consisting of four dimensions, namely (a) mobility, (b) physical activity, (c) social activity, and (d) symptom problem complex. Estimation of QWB index proceeds in the following three steps: (a) assessment of functional status profile, (b) Scaling of responses to derive dimension specific weights for the composite index, and (c) estimation of transition probabilities to derive expected duration in each health state. Construction of synthetic measures like DALY or QALY, already takes into account expected duration in each health state. So measurement of disability weight or its complement quality adjustment weight requires only the first two steps i.e. assessment of functional status and dimension specific weights. Authors of the scale have derived a set of dimension specific weights from valuations by 867 raters and using an equal appearing interval rating procedure (Kaplan, Bush and Berry 1976; 1979; Patrick, Bush and Chen 1973a-b; Blischke, Bush and Kaplan 1975). This instrument has been described by Kaplan, Anderson and Ganiats (1992), and in McDowell and Newell (1987, 1996). QWB was used to gather community valuations for difference health states by the Oregon health services commission (OSHC). The scale was criticised when it was found that weights assigned by it to certain health states were clearly counter intuitive. However, it has been pointed out that the QWB scale was not properly applied by the OSHC and hence the counter intuitive results could not be attributed to the scale.

Local issues in choice of instruments:

It would be useful to recognise certain issues, not covered in the methodological discussions above, and having a bearing on choice of these instruments. I discuss three important issues, namely; (a) need to account for minor and trivial illnesses and its implication for health status measurement, (b) cross cultural validity of health state valuation instruments, and (c) local feasibility and need for adaptation of health state valuation instruments.

Minor and trivial illness:

How to treat the large amount of illnesses that exists but for which people do not seek treatment? It has been observed that aggregate measures like sickness rate (number of persons becoming ill per time period) or illness episodes per time period based on simple count and the implied equal weightage to illnesses of all severity are too stable and non responsive to variations in incidence of more severe illnesses (Logan and Brooke 1957). This is in fact the primary motivation behind the search for a set of unequal health state weights. The relative weight to illness of different severity will depend on the current concept of health. At a time characterised by survival as the dominant concept of health, weights attached to all forms of morbidity are nearly zero. As the concept of health evolves to include absence of disease, more severe forms of disability start receiving higher weightage both in the minds of the patients and public. Further evolution of the concept of health to include quality of life would naturally enhance the weightage received by illnesses considered minor and trivial in an earlier era. Rosser (1983) notes that even though Logan and Brooke, in 1957, sought to increase the sensitivity of aggregate indicators of morbidity by splitting down some categories (there by assigning zero weightage to excluded conditions), such an approach would not relevant in view of increasing concerns about these so called minor and trivial illnesses.

Health status measures differ in sensitivity to minor and trivial illnesses. For example; the Quality of well being scale (QWB) is known to be more sensitive to small departures from perfect health (McDowell and Newell, 1996 p483-491). This is because the QWB construct includes a symptom complex dimension. For example; Erickson et al (1989) found that the QWB scale classified 95% of the 45-64 year old population in less than perfect health compared to 75% when activity limitation was used as the criterion. On the other hand EuroQol is insensitive to small departures from perfect health (McDowell and Newell, 1996 p480-483). This instrument differentiates between more severe forms of morbidity but lumps all morbidities at the healthy end of the scale.

Thus choice of instrument to assign health state weight will have an impact on the kind of policy application to which the resultant aggregate measure of disease burden can be

put. If the health state weighing instrument is not sensitive to minor and trivial illnesses, the resultant disease burden measure could not be used to assess need for ambulatory services.

Cross cultural validity:

Validity of an instrument across different cultural settings will depend on the conception of health from which it arises and the nature of dimensions included in it. If the judgement on some dimensions is influenced by specific cultural characteristics the instrument would not perform well in settings outside the place of its origin. Schumaker and Naughton (1995), recognising the need for portable health status measurement instruments for international use proposed that domains of health related quality of life be restricted to those "universally most essential to one's ability to pursue valued life goals". Most instruments purporting to measure health related quality of life do include such universally useful dimensions such as physical, social and psychological functioning; mobility and self care; and emotional well being. Although this principle is not in doubt, there is scope for wide ranging interpretations of the very generic characterisation of universally useful dimensions.

Another practical issue is ease and accuracy of translation. The translated instrument should retain its original validity and reliability characteristics. Leplege and Verdier (1995) have described methodological aspects of translation of health status measurement instruments. Generic instruments, which have been translated to different languages (Shumaker and Berzon 1995 appendix-1) include (a) the Short form health survey (SF-36); (b) Nottingham health profile (NHP); (c) sickness impact profile (SIP); (d) EuroQol and (e) Dartmouth COOP functional health assessment charts. Out of these five EuroQol is the only instrument to generate a composite index of health status. All other instruments produce general health profiles (McDowell and Newell, 1996).

Local feasibility:

A further issue is feasibility of administration in the local context. In Andhra Pradesh literacy is only about 45%. So the scope for self administration of questionnaires is limited.

Local experience in measurement of health related quality of life is almost non existent. So even for interviewer administered questionnaire the items need to be simple and straight forward to ensure acceptable levels of compliance by interviewers and interviewee.

References:

Allen D., Lee R.H., and Lowson K. 1989. The use of QALYs in health service planning. International Journal of Health Planning and Management 4: 261-73.

Anderson RT, Aaronson NK, Wilkin D. Critical review of the international assessments of health-related quality of life (review). Quality of Life Research 2(6), pp. 369-95. L993.

Bergner M., Bobbit R. A., Carter W. B., and others. 1981. The sickness impact profile: development and final revision of a health status measure. Medical Care 19: 787-805.

Bergner M., Bobbit R. A., Kressel S., and others. 1976. The sickness impact profile: conceptual formulation and methodology for the development of a health status measure. International Journal of Health Services 6: 393-415.

Bergner M., Bobbit R.A., Pollard W. E., and others. 1976a. The sickness impact profile: validation of a health status measure. Medical Care 14: 57-67.

Bleichrodt Han. 1997. Health utility indices and equity considerations. Journal of Health Economics 16: 65-91.

Brooks Richard and EuroQol group. EuroQol: the current state of play. Health Policy. 1996; 37:53-72.

Culyer Anthony J. 1989. The normative economics of health care finance and provision. Oxford Review of Economic Policy 5, no. 1: 34-58.

Euroqol Group. EuroQol. A new facility for measurement of health-related quality of life. Health Policy. 1990; 16:199-208.

Froberg D.G. and Kane R.L. Methodology for measuring health-state preferences, part-I: measurement strategies. Journal of Clinical Epidemiology, 1989; 42(4):345-354.

Froberg D.G. and Kane R.L. Methodology for measuring health-state preferences, part-II: scaling methods. Journal of Clinical Epidemiology. 1989; 42(5):459-471.

Froberg D.G. and Kane R.L. Methodology for measuring health-state preferences, part-III: population and context effects. Journal of Clinical Epidemiology. 1989; 42(6):585-592.

Froberg D.G. and Kane R.L. Methodology for measuring health-state preferences, part-IV: progress and a research agenda. Journal of Clinical Epidemiology.1989; 42(7):675-685.

Gudex C.; Dolan P.; Kind P., and Williams A. Health state valuations from the general public, using the visual analog scale. Quality of Life Research. 1996; 5:521-531.

Huskisson, E. C. (1974). Measurement of pain. Lancet, ii, 1127-31.

Kaplan R. M.; Bush J. W., and Berry C.C. Health Status Index: category rating versus magnitude estimation for measuring levels of well being. Medical Care. 1979; 17:501-523.

Kaplan R. M.; Bush J. W., and Berry C.C. Health status: types of validity and the index of well being. Health Services Research. 1976; 11:478-507.

Kaplan Robert M., Anderson John P., Ganiats Theodore G. The quality of well being scale: rationale for a single quality of life index. in: Walker Stuart R., Rosser Rachel M., Editors. Quality of life assessment. Key issues in the 1990s. Boston//London//Dordrecht: Kluwer Academic Publisher, 1992: 65-94.

Leplege A., and Verdier A. 1995. The adaptation of health status measures: methodological aspects of the translation procedure. in The international assessment of health related quality of life. Theory, translation, measurement and analysis. Editors Shumaker Sally A., and Berzon Richard A.; Oxford, New York: Rapid Communications.

Lodge Milton. Magnitude scaling. Quantitative measurement of opinions. Newburry Park//London//New Delhi: Sage; 1981.

Mas-Collel Andreu, Whinstone Michael D., and Green Jerry. 1995. Micro economic theory. New York, Oxford: Oxford University Press.

McDowell Ian and Newell Claire. Measuring health: a guide to rating scales and questionnaires. New York / Oxford: Oxford University Press; 1987.

McDowell Ian and Newell Claire. Measuring health: a guide to rating scales and questionnaires. New York / Oxford: Oxford University Press; 1996.

McHorney C.A., Ware J. E. Jr., Lu J. F. R., and others. 1994. The MOS 36 item short-form health survey (SF-36): III. Tests of data quality, scaling assumptions, and reliability across diverse patient groups. Medical Care 32: 40-66.

Murray Christopher J.L., and Lopez Alan D. 1996. The global burden of disease in 1990: final results and their sensitivity to alternate epidemiological perspectives, discount rates, age weights and disability weights. in The global burden of disease. A comprehensive assessment of mortality and disability from diseases, injuries, and risk factors in 1990 and projected to 2020. Editors Murray Christopher J. L., and Lopez Alan D. Boston: Harvard School of Public Health.

Nord Erik. The validity of a visual analogue scale in determining social utility weights for health states. International Journal of Health Planning and Management. 1991; 6:234-42.

Nord Erik. The person-trade-off approach to valuing health care programs. Med Decis Making 1995; 15:201-8.

Patrick D. L.; Bush J. W., and Chen M. M. Methods for measuring levels of well-being for health status index. Health Services Research. 1973; 8:228-245.

Pennifer Erickson, Kendall Allen, Anderson John P., and Kaplan Robert M. 1989. Using composite health status measures to assess the Nation's health. Medical Care 27, no. 3 Supplement: S66-S76.

Rosser Rachel. Issues of measurement in the design of health indicators: a review. in: Culyer A.J., Editor. Health indicators. An international study for the European Science Foundation. Oxford: Martin Robertson; 1983; pp. 34-81.

Sackett D.L. and Torrance G.W. The utility of different health states as perceived by general public. Journal of Chronic Diseases. 1978; 7:347-358.

Sen Amartya. Social Choice Theory: A Re-examination. Econometrica 1977; 45:53-90.

Sen Amartya. Personal utilities and public judgements: or What's wrong with Welfare Economics. Economic Journal 1979; 89:537-58.

Shumaker Sally A., Naughton M.J. The international assessment of health related quality of life. in: Shumaker Sally A., Berzon Richard A., Editors. The international assessment of health related quality of life. Theory, translation, measurement and analysis. Oxford//New York: Rapid Communications, 1995.

Shumaker Sally A., and Berzon Richard A; Editors; The international assessment of health related quality of life. Theory, translation, measurement and analysis.; Oxford, New York: Rapid Communications, 1995.

Specter Paul E. 1992. Summated rating scale construction. An introduction. Newburry Park, London, New Delhi: Sage.

Streiner, D. L. & Norman G. R. 1995. Health Measurement Scales: A Practical Guide to their Development and use. New York: Oxford University Press.

Tolley George, Kenkel Donald, and Fabian Robert; 1994; Valuing health for policy. An economic approach; Chicago and London: University of Chicago Press.

Torgerson W.S. 1958. Theory and methods of scaling. New York: John Wiley; page 141. Cited in Patrick et al 1973.

Torrance George W., Sackett David L., Thomas Warren H. Utility maximization model for program evaluation: a demonstration application. in: Berg Robert L., Chairman and Editor. Health status indexes. Proceedings of a conference conducted by Health Services Research, Tucson Arizona, October 1-4, 1972. Chicago: Hospital Research and Educational Trust, 1973.

Torrance George W. Measurement of health state utilities for economic appraisal. A review. Journal of Health Economics. 1986; 5:1-30.

Torrance G.W., and Boyle M. H. 1982. Application of multiattribute utility theory to measure social preferences for health states. Operations Research 30: 1043-69. Cited in Froberg and Kane 1989-I.

Ware J. E., Snow K. K., Kosinski M., and others. 1993. SF-36 health survey: manual and interpretation guide. Boston, MA: The Health Institute, New England Medical Center.

Ware John E. Jr. and Sherbourne Cathy Donald. The MOS 36-item short form health survey (SF-36). Medical Care. 1992 Jun; 30(6).

----------o----------